DOCUMENT RESUME

ED 278 704 TM 870 131

AUTHOR Tollefson, Nona; Chen, Ju Shan
TITLE A Comparison of Item Difficulty and Item Discrimination of Multiple-Choice Items Using "None of the Above" Options.
PUB DATE Oct 86
NOTE 26p.; Paper presented at the Annual Meeting of the Midwest Educational Research Association (Chicago, IL, October, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Difficulty Level; Graduate Students; Higher Education; *Item Analysis; Measurement; *Multiple Choice Tests; Statistics; *Test Format; *Test Items; Vocabulary Skills
IDENTIFIERS Synonyms; Wilcoxon Matched Pairs Signed Ranks Test

ABSTRACT

This study compared item difficulty and item discrimination indices for parallel multiple-choice items in three content areas: measurement concepts, statistical terminology, and synonyms. The statistics and measurement items were administered in classes where graduate students taking the test were studying the content. Vocabulary items represented a common area of knowledge where formal instruction was not offered. Items were written in three formats: a one-correct answer format, a format where "none of the above" was the correct answer, and a format having "none of the above" as a foil. The same pattern of median item difficulties was observed in three experiments. Items having "none of the above" as the correct response had lower median item difficulties than the one-correct answer format and the format using "none of the above" as a foil. The format using "none of the above" as a foil was significantly easier than the format having "none of the above" as the correct answer for measurement concepts and vocabulary items. The same pattern of item difficulty and item discrimination was observed across all item contents. This study concluded that using "none of the above" as a correct answer makes an item more difficult but not necessarily more discriminating. (JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED278704

A Comparison of Item Difficulty and Item Discrimination of Multiple-Choice Items Using "None of the Above" Options

Nona Tollefson and Ju Shan Chen
The University of Kansas

Paper presented at the annual meeting of the Midwest Educational Research Association in Chicago, Illinois, October, 1986.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

N. Tollefson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM 870 131

Multiple-choice items are among the most highly regarded and widely used forms of objective test items (Ebel, 1979, p. 135); and most measurement textbook writers recommend their use over other types of items. Multiple-choice items are favored because they permit extensive sampling of a content domain (Bloom, Madaus, & Hastings, 1981; Brown, 1976) and assessment of the higher levels of the cognitive domain (Anderson, 1972; Bennion, 1977; Brown, 1976; Ebel, 1979; Scannell & Tracy, (1975). Multiple-choice items are adaptable to measurement of most important educational outcomes. They provide a realistic, naturally appropriate setting for testing student achievement and are less susceptible than other forms of recognition items to chance errors resulting from guessing the correct answer (Ebel, 1979).

Disagreement exists among textbook authors about the use of exclusion/inclusion response alternatives such as "none of the above" or "all of the above" in multiple-choice items. Some authors recommend the use of the one-correct answer multiple-choice format (Bennion, 1977; Brown, 1976; Kryspin & Feldhusen, 1974; Mehren & Lehmann, 1978; Scannell & Tracy, 1975), while other authors recommend the use of complex response alternatives (Bloom, Madaus, & Hastings, 1981; Ebel, 1979). Furthermore, the research literature describing the properties (item difficulty and item discrimination) of items with exclusion or inclusion response alternatives is contradictory, and textbook writers give anecdotal reasons rather than empirical data for their recommendations to use or not to use this type of alternative.

Wesman and Bennet (1946) reported no effect on item difficulty or test reliability associated with the "none of the above" alternative on a test administered to nursing students. Boynton (1950) used the "none of the above" option on a clerk and stenographer-typist examination and found

items with the "none of the above" option were more difficult than items using a conventional multiple-choice format. Rimland (1960) used "right answer not given" as an option on the U.S. Navy Arithmetic Test and found that items with this option were more difficult than the one-correct answer items. Williamson and Hopkins (1967) used the "none of these" option on a fourth grade arithmetic test; the form including items using "none of these" as an option had a higher mean test difficulty than a parallel form of the test.

Studies of the effect on item difficulty and discrimination of complex alternatives of the "none of the above," "all of the above," and "A and B are correct" form have yielded contradictory results. Hughes and Trimble (1965) compared the single-correct answer format to the "both 1 and 2 are correct," the "all of the above," and "none of the above" formats on a college-level psychology course. These researchers found the mean score for the test form including inclusion/exclusion alternatives to be significantly lower than the mean score for the test form including items having only one-correct answer.

Mueller (1975) compared conventional multiple-choice items to multiple-choice items with three different types of complex alternatives: "all of the above," "none of the above," and combinations of alternatives such as "A and B are correct". The items were included on an examination for a real estate course. Mueller concluded that the format using the combination of alternatives was the most difficult.

Tollefson and Tripp (1983) compared the item difficulty and item discrimination of three multiple-choice item formats, two complex alternative formats, and a one-correct answer format. Items having "none of the above" as the correct answer had a significantly higher mean discrimination index

than the other two formats. The "none of the above" as the correct answer items were also more difficult than the other forms. Oosterhof and Coats (1984) also found items using "none of the above" as the correct answer to be more difficult.

In a recent study, Tollefson (in press) compared multiple-choice items written in three formats: one-correct answer, "none of the above" as a foil, and "none of the above" as the correct answer. The format using "none of the above" as the correct answer was the most difficult of the three formats. No consistent differences in item discrimination indices were found among the three formats. The test form composed of multiple-choice items with one-correct answer was more reliable than the two test forms using the "none of the above" option.

Purpose of the Study

The present study compared item difficulty and item discrimination indices for parallel multiple-choice items in three content areas: measurement concepts, statistical terminology, and synonyms. These content areas are representative of verbal and quantitative subject-matter fields.

The three content areas selected also permitted comparison of item characteristics in tests measuring instructed and uninstructed material. The statistics and measurement items were administered in classes where students taking the test were studying the content. Vocabulary items represented a common area of knowledge, but an area where formal instruction is not offered.

Method

Three experiments were conducted. The first used statistical concepts; the second used measurement concepts; and the third used vocabulary items as stimulus materials. The methodology in each experiment was to include a set of experimental items in a longer test. All tests were administered under testing conditions.

Items in the experimental set were written in three formats: a one-correct answer format, a format where "none of the above" was the correct answer, and a format having "none of the above" as a foil. The stem of the measurement and statistical items were written as questions to insure that each item contained a complete idea in the stem. The "none of the above" option always appeared as option "d". When "none of the above" was the correct response, it replaced the correct answer. When "none of the above" was a foil, it replaced the least frequently chosen option in the one-correct answer format. All items had been field-tested prior to their inclusion in the experimental set of items.

Vocabulary items were selected to be of moderate difficulty. The difficulty of the items was controlled by selecting either moderately difficult stimulus items or moderately difficult options. For example, a relatively easy stimulus item such as frivolous was made more difficult by use of the synonym paltry as the correct answer.

In experiment 1, all items using the same format were assigned to one form of the test. For experiment 2 and experiment 3, items were randomly assigned to form and all forms included all item formats.

Experiment 1

Subjects (n = 78) in this study were graduate students enrolled in two sections of a first year statistics class. Students were randomly assigned to complete one of three forms of a 73-item final examination in the course. Twelve items were included in the experimental set of items. Experimental items were written in three formats: a one-correct answer format (Form C), a format where "none of the above" was the correct answer (Form B), and a format where "none of the above was a foil (Form A). All items having the same format (i.e., "none of the above" as a foil) were assigned to one form of the test in this experiment. To avoid a response set, all forms of the test included a common set of items where "all of the above" and "none of the above" were foils and the correct answer. The use of a common set of items where "none of the above" was the correct answer and a foil eliminated the possible problem of "none of the above" being the correct answer everytime this option appeared in an item. Figure 1 presents one of the experimental items.

Insert Figure 1 about here

Item difficulty (proportion of students answering the item correctly) and item discrimination indexes (the point-biserial correlation between the item and the total score) were computed for each of the experimental items. The distribution of item difficulties for the statistical concepts items was negatively skewed. The median item difficulties ranged from .86 for items where "none of the above" was a foil to .65 where "none of the above" was the correct answer. Item discrimination values ranged from -.03 to .90

and most (58%) were .30 or higher. Table 1 reports the item difficulty values; Table 2 gives the item discrimination values.

Insert Tables 1 and 2 about here

The Wilcoxon matched-pairs signed-ranks test was used to test the hypothesis that the population distributions of item difficulties were identical for the three item formats. These tests indicated a significant difference in distributions for items with one-correct answer and items having "none of the above" as the correct answer. Items written in the one-correct answer format were easier than items having "none of the above" as the correct answer ($T(12) = 2.5$, $p < .05$). Comparisons of one-correct answer items and items having "none of the above" as a foil and items using "none of the above" as the correct answer and as a foil yielded nonsignificant T values.

The highest average (median) discrimination value was observed for items with one-correct response. However, the matched-pairs signed-ranks tests indicated nonsignificant T values for the comparisons of item discrimination values.

Experiment 2

Subjects in this study were graduate students enrolled in a graduate measurement class ($n = 54$). Students were randomly assigned to one of three forms of a 24-item test ($n = 19$ for Form A, $n = 18$ for Form B, and $n = 17$ for Form C). This test included 11 experimental items written in the same three formats described in experiment 1. However, in this experiment, each form included all item formats. Figure 2 shows the assignment of item formats to test forms. Figure 3 presents a sample item.

Insert Figures 2 and 3 about here

Item difficulty and item discrimination values were computed for each item. Item difficulty indices for the measurement items were negatively skewed. Most items had difficulty indexes of .80 or higher. Item discrimination values ranged from -.21 to .76 and most items had discrimination values of .30 or higher.

Wilcoxon matched-pairs signed-ranks tests were used to test for differences in population distributions. Tables 3 and 4 report the item difficulties, item discriminations, and T values for these analyses. Items written in the one-correct response format were significantly easier than items where "none of the above" was the correct response ($T(11) = 2$, $p < .05$). Items having "none of the above" as the correct response were significantly more difficult than parallel items having "none of the above" as a foil ($T(8) = 4$, $p < .05$). Analysis of discrimination values for this set of items yielded no significant results.

Insert Tables 3 and 4 about here

Experiment 3

In this experiment, 52 graduate students enrolled in a teaching methods class were randomly assigned to complete one of three forms of a 40-item vocabulary test. In this experiment, 18 items were included in the experimental set and written in the three formats. The items were randomly assigned to form and all item formats were included in all forms of the test. Figure 4 shows the assignment of item formats to test forms. Figure 5 illustrates the type of vocabulary items used.

Insert Figures 4 and 5 about here

Item difficulty and item discrimination indexes were computed for each item and are reported in Tables 5 and 6. Vocabulary items were more difficult than the items mesuring statistical and measurement concepts. Median item difficulties ranged from .54 for items with one-correct response to .37 for item with "none of the above" as the correct answer. The median item discrimination values were comparable to those for items measuring statistical and measurement concepts. In this experiment, the same pattern of item difficulties were observed across the three formats. Items with one-correct answer were, on average, easier than matched items using the "none of the above" option; and the set of items using "none of the above" as a correct answer was significantly more difficult than the set using "none of the above" as a foil (T (18) = 29.5, $p < .05$). Wilcoxon matched-pairs signed-ranks tests indicated no significant differences in the distribution of item discrimination values for the set of vocabulary items.

Insert Tables 5 and 6 about here

Discussion

This study compared the effect on item difficulty and item discrimination of using an exclusion alternative, "none of the above," as the correct answer and as a foil. Across all experiments, the same pattern of median item difficulties was observed. Experimental items having "none of the

above" as the correct response had lower median item difficulties than the one-correct answer format and the format using "none of the above" as a foil. The format using "none of the above" as a foil had a lower median item difficulty than the one-correct answer format. The one-correct answer format was significantly easier than the format having "none of the above" as the correct answer for items measuring statistical and measurement concepts. The format using "none of the above" as a foil was significantly easier than the format having "none of the above" as the correct answer for measurement concepts and vocabulary items. In general, the median item discrimination values were higher for the one-correct answer format than they were for the exclusion formats. However, no consistent pattern of significant differences in item discrimination values across the three item formats was observed.

The same pattern of item difficulty and item discrimination was observed across all item contents. Items with "none of the above" as the correct answer were the most difficult for vocabulary items; however, this item format remained the most difficult in relative terms for the measurement and statistics items. This finding suggests that instruction in the content does not change the relative difficulty of items using "none of the above" as the correct answer.

The method used to assign item formats to test form did not influence the pattern of item difficulties. Whether all items having the same format were assigned to the same test form (experiment 1) or randomly assigned to different test forms, the pattern of item difficulties remained the same.

Guidelines for writing multiple-choice items suggest that the item writer use a one-correct answer format. The results of this study suggest that this format produces items that are easier and more discriminating

than items using exclusion formats. The preference for use of an exclusion alternative is strong in content areas requiring the solution of mathematical problems. Oosterhof and Coats (1984) used this option with business finance problems and found it to be more difficult than the other options studied. This study did not use mathematical problems but found that the "none of the above" option as the correct answer was consistently more difficult than a one-correct answer fo[illegible]hree differe[illegible] content areas.

The difficulty posed by an exclusion optio[illegible]t from the fact that it requires the respondent to assume a set that is opposite from the manner in which instruction is typically offered. We tend to teach "what is" rather than "what is not." Furthermore, for any single item, the set of "what is" is probably much smaller than the set of "what is not." It is also possible that recognizing "none of the above" is the correct response may reflect greater understanding of the material than recognizing one answer is correct. Combs (1953) wrote: "Complete knowledge, knowing what is right, is equivalent to knowing everything that is wrong (p. 308). So, it may be that using an exclusion option as the correct answer measures more complete understanding of the material tested. If this were true, however, one would expect items using "none of the above" as the correct response to be more discriminating than other item formats. This was not the case in this study.

Tests provide a means of assessing the respondent's level of information or misinformation and it is important that we use item formats that best enable us to accomplish this objective. Therefore, we need to continue to investigate the properties of the several forms of multiple-choice items. This study and related studies suggest that using "none of the

above" as a correct answer makes an item more difficult but not necessarily more discriminating. Researchers need to continue to explore the properties of multiplle-choice items using exclusion or inclusion options as the correct answer. Information that increases knowledge of the strengths and limitations of multiple-choice items will add to our skill in measuring student achievement.

References

Anderson, R. C. (1972). How to construct achievement tests to assess comprehension. Review of Educational Research, 42, 145-170.

Bennion, D. H. (1977). Assessing student learning: Behavioral objectives and evaluation procedures for the affective, cognitive and psychomotor domains. Dubuque: Kendall-Hunt.

Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). Evaluation to improve learning. New York: McGraw-Hill.

Boynton, M. (1950). Inclusion of 'none of these' makes spelling items more difficult. Educational and Psychological Measurement, 10, 431-432.

Brown, F. A. (1976). Principles of educational and psychological testing (2nd ed.). New York: Holt, Rinehart & Winston.

Coombs, C. H. (1953). On the use of objective examinations. Educational and Psychological Measurement, 13, 308-310.

Ebel, R. L. (1979). Essentials of educational measurement (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Hughes, H. H., & Trimble, W. E. (1965). The use of complex alternatives in multiple-choice items. Educational and Psychological Measurement, 25, 117-126.

Kryspin, W. J., & Feldhusen, J. F. (19740. Developing classroom tests: A guide for writing and evaluting test items. Minneapolis: Burgess.

Mehren, W. A., & Lehmann, I. J. (1978). Measurement and evaluation in education and psychology (2nd ed.). New York: Holt, Rinehart & Winston.

Mueller, D. J. (1975). An assessment of the effectiveness of complex alternatives in multiple-choice achievement test items. Educational and Psychological Measurement, 35, 135-141.

Oosterhof, A. C., & Coats, P. K. (1984). Comparison of difficulties and reliabilities of quantitative word problems in completion and multiple-choice item formats. Applied Psychological Measurement, 8, 287-294.

Rimland, B. (1960). The effect of varying time limits and use of 'right answer not given' in experiment forms of the U.S. Navy Arithmetic test. Educational and Psychological Measurement, 20, 533-539.

Scannell, D. P., & Tracy, D. B. (1975). Testing and measurement in the classroom. Boston: Houghton Miffling.

Tollefson, Nona & Tripp, A. (1983). The effect of item format on item difficulty and item discrimination. Paper presented at the annual meeting of the American Educational Research Association. Montreal, Canada.

Tollefson, N. (In Press). A comparison of the item difficulty and item discrimination of multiple-choice items using "none of the above" and one correct response option. Educational and Psychological Measurement.

Weiten, W. (1982). Relative effectiveness of single and double multiple-choice questions in educational measurement. Journal of Educational Research, 5, 46-50.

Wesman, A. G., & Bennett, G. K. (1946). The use of 'none of these' as an option in test construction. Journal of Educational Psychology, 27, 541-549.

Williamson, M. L., & Hopkins, K. D. (1967). The use of 'none of these' versus homogeneous alternatives on multiple-choice tests: Experimental reliability and validity comparisons. Journal of Educational Measurements, 4, 53-57.

Figure 1
Sample of an Item Included in the Statistical Concepts Test

Item Stem: In which of the following situations would a t test for two independent samples be an appropriate statistic to use?

One Correct Answer:

a. determining if the median income for male and female faculty was equal
b. evaluating the degree of association between years in school and student perceptions of effective teachers
c. judging which of two methods was more effective at increasing achievement test scores for two groups of students matched on age, sex, and ability
d. determining if there is a difference in end of course grades between students who were randomly assigned to receive a new textbook and those randomly assigned to use the old text book.

"None of the Above" as the correct answer:

a. determining if the median income for male and female faculty was equal
b. evaluating the degree of association between years in school and student perceptions of effective teachers
c. judging which of two methods was more effective at increasing achievement test scores for two groups of students matched on age, sex, and ability
d. none of the above

"None of the Above" as a foil:

a. determining if the median income for male and female faculty was equal
b. evaluating the degree of association between years in school and student perceptions of effective teachers
c. judging which of two methods was more effective at increasing achievement test scores for two groups of students matched on age, sex, and ability
d. none of the above

Figure 2
Assignment of Items to Test Form for Experiment 2, Measurement Items

Form A	Form B	Form C
2	3	1
2	3	1
3	1	2
1	2	3
1	2	3
1	2	3
2	3	1
3	1	2
1	2	3
3	1	2
1	2	3
n = 19	18	17

1 = one-correct response

2 = "none-of-the-above" as the correct answer

3 = "none-of-the-above" as a foil

n = number of people completing each form

Figure 3
A Sample Item From the Measurement Test

Item Stem: Changes in an individual's responses to a test as a result of learning contribute to error variance for which type of coefficient?

One Correct Answer

A. coefficient of stability
B. coefficient of equivalence
C. coefficient of homogeneity
D. coefficient of internal consistency

"None of the Above" as the correct answer

A. coefficient of homogeneity
B. coefficient of equivalence
C. coefficient of internal consistency
D. none of the above

"None of the Above" as a foil

A. coefficient of stability
B. coefficient of equivalence
C. coefficient of internal consistency
D. none of the above

Figure 4
Assignment of Item Formats to Test Forms for the Vocabulary Items

Form A	Form B	Form C
2	3	1
1	2	3
1	2	3
1	2	3
2	3	1
3	1	1
1	2	3
1	2	3
3	1	2
2	3	1
2	3	1
2	3	1
2	1	3
2	1	3
1	2	3
3	1	2
1	2	3
3	1	2
n = 28	26	25

1 = one-correct response

2 = "none of the above" as the correct response

3 = "none of the above" as a foil

n = number of people completing each form

Figure 5
Sample Vocabulary Item From Experiment 3

Item: For each word, select the word that means the same thing or most nearly the same thing as the underlined word.

defile

One Correct Answer

A. vitiate
B. debauch
C. sanctify
D. hallow

"None of the Above" as the correct response

A. vitiate
B. sanctify
C. hallow
D. none of the above

"None of the Above" as a foil

A. vitiate
B. sanctify
C. debauch
D. none of the above

Table 1
Item Difficulty Indices for Items Measuring Statistical Concepts

	Form C One Correct Response	Form B "None of the Above" Correct	Form A "None of the Above" Foil
	.81	.65	.60
	.92	.65	.96
	.86	.60	.93
	1.00	.92	1.00
	.81	.56	.87
	.89	.28	.88
	.78	.69	.35
	.90	.89	.63
	.91	.54	.88
	.63	.32	.38
	.80	.84	.85
	.68	.64	.42
Mdn	.83	.65	.86
N	26	25	27

A vs B $n = 12$ $T = 2.5$ $p < .05$
A vs C $n = 12$ $\bar{T} = 18$
B vs C $n = 12$ $\bar{T} = 40$

N indicates the number of people completing each form

n indicates the number of items in the experimental set of items

Table 2
Item Discrimination Indices for Items Measuring Statistical Concepts

	Form C One Correct Response	Form B "None of the Above" Correct	Form A "None of the Above" Foil
	.90	.50	.34
	.60	.50	.33
	.67	.22	.48
	.00	.09	.00
	.55	.57	-.03
	.21	.22	.38
	.79	.32	.48
	.29	.29	.53
	.19	.46	.49
	.05	.12	.27
	.56	.62	.42
	.70	.23	.21
Mdn	.55	.31	.33
N	26	25	27

A vs B, n = 12 T = 43
A vs C, n = 12 T = 18
B vs C n = 12 T = 39

N indicates the number of people completing each form

n indicates the number of items in the experimental set of items

Table 3
Difficulty Indices for the Measurement Items

	One Correct Response	"None of the Above" Correct	"None of the Above" Foil
	.94	.53	1.00
	.53	.47	.78
	1.00	.53	.84
	1.00	.89	.94
	.74	.83	.82
	.89	.06	.94
	1.00	1.00	1.00
	.83	.35	.58
	.95	.67	.59
	.72	.65	.65
	.84	.94	.94
Mdn	.89	.65	.84

A vs B $n = 11$ $T = 2$ $p < .05$
A vs C $n = 10$ $T = 17$
C vs B $n = 8$ $T = 4$ $p < .04$

n indicates the number of items in the experimental set of items. ns differ across comparisons because items with identical difficulties are eliminated from the analysis.

Table 4
Discrimination Indices for the Measurement Items

One Correct Response	"None of the Above" Correct	"None of the Above" Foil
-.21	.31	0.00
.62	.33	.67
0.00	.27	.71
0.00	.11	.62
.18	.47	.33
.26	.21	-.13
0.00	0.00	0.00
.67	.09	.42
.58	.43	.33
.43	.49	.49
.76	.53	.29
Mdn .26	.31	.33

A vs B $n = 10$ $T = 30.5$
A vs C $n = 10$ $T = 26$

n indicates the number of items in the experimental set of items. ns differ across comparisons because items with identical discrimination indexes are eliminated from the data analysis.

Table 5
Item Difficulty Indexes for the Vocabulary Items

Item	One Correct Answer	"None of the Above" Correct Answer	"None of the Above" a Foil
	.80	.96	.88
	.61	.58	.44
	.61	.08	.36
	.50	.77	.24
	.68	.61	.62
	.42	.44	.29
	.86	.31	.08
	.29	.31	.28
	.81	.08	.57
	.56	.21	.54
	.76	.29	.50
	.28	.29	.27
	.31	.46	.44
	.54	.57	.60
	.43	.23	.52
	.50	.64	.34
	.61	.65	.28
	.23	.28	.21
Mdn	.54	.37	.40

A vs B	$n = 18$	$T = 56.5$	
A vs C	$n = 18$	$\bar{T} = 29.5$	$p < .05$
C vs B	$n = 18$	$\underline{T} = 72.5$	

n indicates the number of items in the experimental set of items.

Table 6
Item Discrimination Indexes for Vocabulary Items

One Correct Answer	"None of the Above" Correct Answer	"None of the Above" a Foil
.20	.40	.04
.53	.16	.09
.21	.12	.50
.07	.36	.36
.48	.45	.15
.71	.58	.45
.74	-.11	-.36
.53	.50	.53
.36	.45	.25
.37	.05	.23
.59	.43	.27
.40	.33	-.25
.36	.26	.39
.30	.36	.43
-.02	-.22	.30
.55	.26	.29
.26	.66	.11
.17	-.09	.48
Mdn 36.5	34.5	.28

A vs B $n = 18$ $T = 63.5$
B vs B $n = 17$ $\overline{T} = 46.5$
C vs B $n = 17$ $T = 80$

n indicates the number of items in the experimental set of items, ns differ across comaprisons because items with identical discrimination indexes are eliminated from the analysis.